Twenty-four near-instabilities of Caspar-Klug viruses

François Englert,[1] Kasper Peeters,[2] and Anne Taormina[3]
[1]*Service de Physique Théorique and The International Solvay Institutes, Université Libre de Bruxelles, Belgium*
[2]*Institute for Theoretical Physics, Utrecht University, The Netherlands*
[3]*Department of Mathematical Sciences, Durham University, United Kingdom*
(Received 27 April 2008; published 10 September 2008)

Group theoretical arguments combined with normal mode analysis techniques are applied to a coarse-grained approximation of icosahedral viral capsids which incorporates areas of variable flexibility. This highlights a remarkable structure of the low-frequency spectrum in this approximation, namely, the existence of a plateau of 24 near zero modes with universal group theory content.

DOI: 10.1103/PhysRevE.78.031908 PACS number(s): 87.15.ad, 87.15.M−, 89.75.Fb

I. INTRODUCTION AND SUMMARY

Proteins in a thermal bath exhibit a wide spectrum of dynamical behaviors which can be probed, for instance, by inelastic neutron diffusion (see Ref. [1] for a review). In particular, they undergo slow, large-amplitude motions which are now widely believed to be instrumental to their function. These ideas were first tested in Refs. [2–7], where normal mode analysis (NMA) was used to argue that only a few low-frequency normal modes of vibration are sufficient to describe, with great accuracy, the conformational changes in a variety of proteins. The method has limitations, as it assumes the existence of a single potential well whose minimum is a given stable conformation of the protein under study, and therefore overlooks the possibility of neighboring multiminima of energy reported to exist in Refs. [8–10]. Yet, this technique yields dynamical data which are consistent with experimental results on proteins, as observed on case-by-case studies in Refs. [11–14], and is also confirmed by a recent statistical analysis [15]. Although biologically significant low-frequency motions are typically not vibrational due to the damping influence of the protein environment, NMA captures the tendency of the biomolecule to change in a few particular directions corresponding to low-frequency normal modes, and thus remains a useful tool when time-dependent methods such as molecular dynamics are prohibitive.

The success of NMA in studying protein dynamics has prompted its use in the context of large macrobiomolecular assemblies. The main motivation so far has been to pin down whether several experimentally observed conformations of viral particles could be inferred from one another by arguing that conformational changes occur in directions which maximally overlap with those of a few low-frequency putative normal modes of vibration of the capsid [16–19]. The biggest challenge remains the choice, within the NMA framework, of a potential which optimally captures the physics of capsid vibrations while taking into account a reduced number of degrees of freedom to enable practical calculations. Many NMA applied to viruses implement variations of the simple elastic network model proposed a decade ago [20], in which the atoms are taken as point masses connected by springs modeling interatomic forces, provided the distance between them is smaller than a given cutoff parameter. Simplified versions include the restriction to C^{α} atoms only, the approximation in which each residue is considered as a point mass, or where even larger domains within the constituent coat proteins are treated as rigid blocks [17]. In an effort to optimize the NMA techniques when applied to particles with high symmetry, group theoretical considerations have also been exploited [16,18]. When combined with an elastic network approach, it allows for more extensive normal modes calculations of viral capsids [18] and compares well with results obtained using an elastic network-RTB setup [17].

The elastic potential in all analyses above has two major drawbacks: It does not discriminate between strong and weak bonds as it depends on a single spring constant and it uses the rather crude technique of increasing the distance cutoff to resolve capsid instabilities. Consequently, the frequency spectra have much less structure than one would expect in reality, and in particular fail to reproduce areas of rigidity and flexibility of the capsid satisfactorily. This problem has been addressed in [21] where a bond-cutoff method is implemented, together with different spring constants for the various types of chemical interactions. The proposed model reproduces conformational changes better than the conventional distance-cutoff simulations.

In this paper, we propose a reductionist approach to the study of icosahedral viral capsid vibrations within the harmonic approximation. We consider Caspar-Klug viral capsids, which are classified according to a triangulation number T [22]. A $T=n$ capsid [28] exhibits $60n$ coat proteins, organized in clusters of twelve pentamers located at the vertices of an icosahedron, and $10(n-1)$ hexamers at global threefold and/or local sixfold symmetry axes of the icosahedral capsid. In our coarse-graining, these coat proteins are approximated by point masses located at their centers of mass, calculated using the Protein Database (PDB). We set up an elastic network whose representatives are these point masses, with all masses normalized to 1. The network topology is determined by data from the VIPER website [23]: two point masses are connected by a spring whenever VIPER provides a value for the association energy of the two corresponding proteins. The spring constants of our model are of the form $\kappa_{mn}=\rho_{mn}\kappa$, with ρ_{mn} the ratio of the association energy of protein pair (m,n) to the largest association energy listed in VIPER, and κ is a free parameter which reflects the lack of confidence in the absolute values of association energies published in VIPER. The potential reads

$$V = \sum_{\substack{m<n \\ m,n=1}}^{N} \frac{1}{2}\kappa_{mn}(|\vec{r}_m - \vec{r}_n| - |\vec{r}_m^0 - \vec{r}_n^0|)^2, \quad (1)$$

where the vector $\vec{r}_m^0$ refers to the equilibrium position of protein m and the vector $\vec{r}_m$ to its position after elastic displacement, all vectors originating from the center of the capsid. The resulting force (stiffness) matrix thus depends on the parameter κ, and possesses more than the six zero eigenvalues expected from the rotations and translations of the whole capsid whenever the number of constraints imposed by the association energies is smaller than $180n-6$, where $180n$ is the total number of degrees of freedom for a $T=n$ capsid.

Our main result, which will be substantiated by group theory arguments below, is the existence, in the low-frequency spectrum of all our coarse-grained stable Caspar-Klug capsids, of 24 near-zero normal modes of vibration which always fall in the same set of nonsinglet irreducible representations of the icosahedral group [see Eq. (4)]. The first singlet representation, which is associated with a fully symmetric mode, always appears higher up in the spectrum, in accordance with the expectation that such a motion requires more energy to develop. The presence of 24 near-zero modes in the spectrum of viral capsids is deeply rooted in the fact that the latter exhibit icosahedral symmetry. Although our model is too crude to account for a realistic flexibility of the viral capsid, and therefore too primitive to get quantitative information on the virus function, it can serve as a first step in an analysis which should bridge the gap with existing molecular dynamics simulations. It discriminates between strong and weak interprotein bonds and therefore captures effects of varying flexibility across the capsid. Moreover, it is simple enough to highlight universal aspects of the low-frequency spectrum of normal modes of vibration, inherited from the underlying icosahedral symmetry, and to pin down the influence of the elastic network design on the instability of the capsid.

Our results should thus be viewed as a remarkable property of the lowest order approximation to virus capsid vibrations, which may potentially be relevant once subsequent orders, with additional interactions and additional degrees of freedom, are taken into account. A meaningful comparison with existing virus NMA computations in the literature is difficult at this stage, partly because of reasons highlighted in Ref. [21] regarding the use of Tirion's potential and the distance-cutoff method. However, it is noteworthy that the first nonzero mode (a singlet of $\mathcal{I}$) in the spectrum of HK97, as calculated in Ref. [19] using an all-atom simulation with a cutoff, appears at mode 31 and is therefore compatible with our qualitative arguments on the number of low-frequency modes [29].

II. LOW-FREQUENCY PLATEAU

The remainder of this paper provides the group theoretical arguments leading to the 24 near zero-mode plateau in our coarse-grained capsids. We first approximate a Caspar-Klug $T=n$ capsid by merging the $3n$ proteins per icosahedral face into a single point mass whose equilibrium position is at the center of the face. These point masses are connected by springs (masses and spring constants normalized to 1) between nearest neighbors, forming a dodecahedral cage with 30 edges, dual to the icosahedron. Such a structure is unstable, since the number of genuine degrees of freedom is $3\times 20-6=54$, while there are only 30 constraints. Accordingly, the cage develops 24 nontrivial zero modes. To analyze these instabilities, we compare all possible motions of this dodecahedral cage consistent with icosahedral symmetry, with the dodecahedral motions induced by all possible motions of the 12 vertices of an icosahedron. By induced, we mean that the dual dodecahedron moves in such a way that its vertices are located at the centers of the (deformed) icosahedral faces at all times.

Standard group theoretical methods reviewed in Ref. [24] are used to calculate the decomposition of the 36-dimensional displacement representation $\Gamma_{\mathrm{ICO}}^{\mathrm{displ},36}$ of the vertices into irreducible representations of the full icosahedral group $\mathcal{I}_h$, which contains 60 proper rotations and an extra 60 elements obtained by multiplication of the latter by the inversion operation. The result is (see also Ref. [25]),

$$\Gamma_{\mathrm{ICO}}^{\mathrm{displ},36} = \Gamma_+^1 + \Gamma_+^3 + 2\Gamma_-^3 + \Gamma_-^{3'} + \Gamma_+^4 + \Gamma_-^4 + 2\Gamma_+^5 + \Gamma_-^5, \quad (2)$$

where the numerical superscripts indicate the dimensionality of the corresponding irreducible representation, while the $\pm$ subscripts refer to different irreducible representations of same dimension. If the subgroup $\mathcal{I}$ of 60 proper rotations is used instead, $\pm$ representations are indistinguishable. The same group theoretical method yields, for the motion of the dodecahedral cage, the 60-dimensional displacement representation $\Gamma_{\mathrm{DODE}}^{\mathrm{displ},60}$,

$$\Gamma_{\mathrm{DODE}}^{\mathrm{displ},60} = \Gamma_+^1 + \Gamma_+^3 + 2\Gamma_-^3 + \Gamma_-^{3'} + \Gamma_+^4 + \Gamma_-^4 + 2\Gamma_+^5 + \Gamma_-^5 + \Gamma_+^{3'} + \Gamma_-^{3'} + \Gamma_+^4 + \Gamma_-^4 + \Gamma_+^5 + \Gamma_-^5. \quad (3)$$

Although Eq. (2) depends only on icosahedral symmetry and not on actual links between icosahedral vertices, it is useful to visualize an icosahedral cage formed by point masses at its vertices joined by identical springs along its 30 edges. In contradistinction with the dodecahedral cage, the icosahedral cage would have no nontrivial zero modes, in accordance with the fact that the number of its genuine degrees of freedom $3\times 12-6=30$ is equal to the number of its constraints. The spectra of the icosahedral and dodecahedral cages are depicted in Figs. 1(a) and 1(b).

As no icosahedral motion leaves the dual dodecahedron fixed, the icosahedral motions induce a 36-dimensional vector space of (infinitesimal) motions of the dual dodecahedron. These include the six global zero modes $\Gamma_+^3+\Gamma_-^3$ which are identical in both systems (as well as the global dilation vibrational mode Γ_+^1). All icosahedral normal modes of vibration pertaining to an irreducible representation of $\mathcal{I}_h$ in $\Gamma_{\mathrm{ICO}}^{\mathrm{displ},36}$ must be linear combinations of dodecahedral normal modes belonging to the same irreducible representation. Hence, provided that the vibrational modes of the icosahedron have nonvanishing components in the vibrational modes of finite frequency of the dodecahedron, as is true by inspection, the set of representations in Eq. (3) which are not contained in Eq. (2) describe the nontrivial zero modes of the

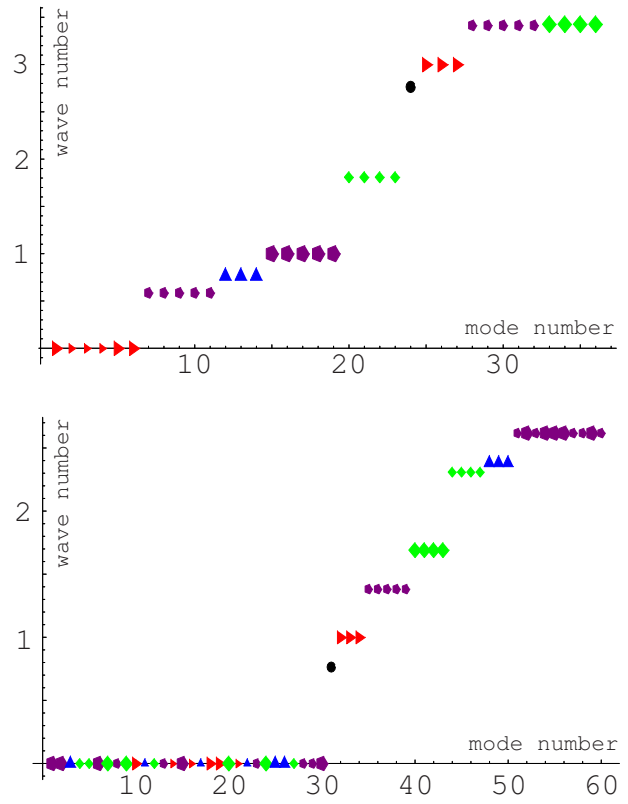

FIG. 1. (Color online) Frequencies of normal modes of vibration for (a) an icosahedral cage, (b) a dodecahedral cage. The modes ▶, respectively, ▲, belong to three-dimensional irreducible representations $\Gamma_{\pm}^{3}$ respectively, $\Gamma_{\pm}^{3'}$ of the icosahedral group. The diamond (pentagon) modes belong to four- (five-) dimensional irreducible representations. Small (large) symbols refer to even (odd) parity. The x axis labels the normal modes while the y axis gives the wave numbers up to an overall normalization.

dodecahedron. It follows that a capsid whose proteins would be modeled by 20 point masses located at the center of each face of an icosahedron, and with 30 springs organized in a network modeled at equilibrium by a dodecahedron, develops 24 zero modes of vibration which are organized in the following irreducible representations of the icosahedral group:

$$\Gamma_{+}^{3'} + \Gamma_{-}^{3'} + \Gamma_{+}^{4} + \Gamma_{-}^{4} + \Gamma_{+}^{5} + \Gamma_{-}^{5}. \tag{4}$$

Note that the zero modes belonging to the $\Gamma_{+}^{3'}$ representation appear in Eq. (3) but have no counterpart in Eq. (2). Hence they must induce no motion when mapped back to an icosahedral system, that is, the sum of the displacements of any five vertices defining a dodecahedral face must be zero. This is indeed the case as illustrated in Fig. 2, where the icosahedral system may be thought of as constructed from the center-of-mass positions of the dodecahedron vertices of each pentagon face. The stabilization of such a capsid requires the introduction of at least 24 further springs in a manner that respects the icosahedral symmetry. The magnitudes of the spring constants determine how much the zero modes are lifted from zero, and shape the structure of the created low-frequency plateau.

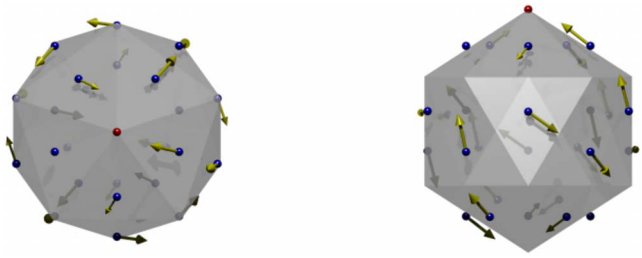

FIG. 2. (Color online) Top and side views of the normal modes of a dodecahedral system belonging to the irreducible representation $\Gamma^{3'+}$ of the icosahedral group $\mathcal{I}$.

Despite its simplicity, the approximation just described catches an essential feature of the Caspar-Klug viruses, namely, the 24-state low-energy plateau with group content (4). Indeed, consider again a $T=n$ capsid with $3n$ proteins per icosahedral face but now treat them, in accordance with our introductory discussions, as $3n$ point masses located at the corresponding protein's center of mass calculated from the PDB files. We link them by springs according to the association energies listed in VIPER [30]. Neglecting first the links between different faces, we would expect a number of capsid zero modes, $N_0=(9n-3k)20=60(3n-k)$. Here the $9n$ degrees of freedom within a face are constrained only by $3k$ links, k an integer, as a consequence of the global threefold symmetry of the icosahedron whose axis passes through the center of the icosahedral face. A stable capsid is characterized by a force matrix having exactly six zero modes, so that, in principle, one needs to introduce at least $60(3n-k)-6$ independent constraints (i.e., bonds) to stabilize the whole capsid. How these must be chosen in a three-dimensional context is a mathematical question which requires further investigation. It is, however, interesting to note that the VIPER website gives association energies for some interface proteins, which correspond to "[icosahedron] edge-crossing" bonds. By symmetry, these come as multiples of 30 or 60 when considering the capsid as a whole. We thus expect that, keeping the $2(3n-k)-1$ edge-crossing bonds per edge with the largest association energies, would yield a capsid with 24 nontrivial zero modes. In all stable coarse-grained capsids we studied [26], this indeed happens, and adding extra edge-crossing bonds lifts the 24 zero modes appearing in Eq. (4) to a low-frequency plateau. Our earlier approximation thus simply amounts to replacing $3n-k$ by unity in the above analysis.

III. EXAMPLES

We now illustrate our considerations with two examples. The VIPER data for the $T=3$ Rice Yellow Mottle Virus (PDB: 1f2n) is sufficient to stabilize the capsid. It is reproduced schematically in Fig. 3 together with the 40 lowest-frequency normal modes of vibration. There are actually 11 bonds in Fig. 3, but 3 of them are superimposed [two linking green C chains, and one linking a C and a B chain (red)]. The total number of bonds is obtained from the figure using the global twofold, threefold, and fivefold symmetries of the

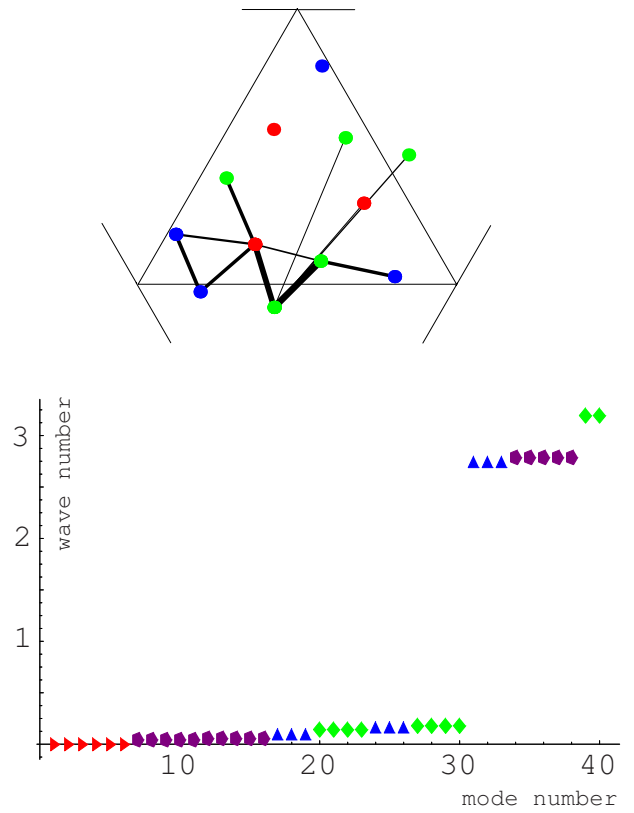

FIG. 3. (Color online) (a) Bond structure of RYMV as given in VIPER. (b) Its forty lowest frequency modes showing a plateau of 24 near-zero modes. Symbols and axes as in Fig. 1.

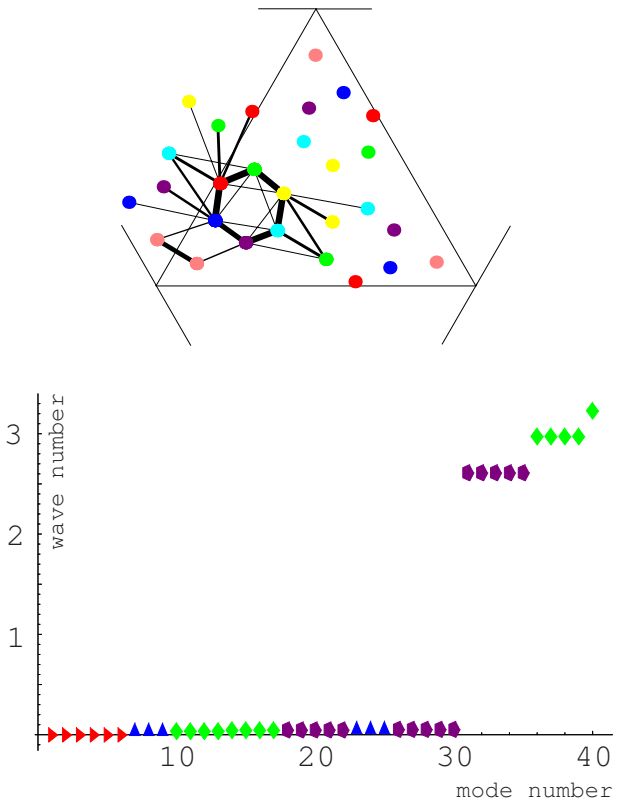

FIG. 4. (Color online) (a) Bond structure of HK97 as given in VIPER. (b) Its forty lowest frequency modes showing a plateau of 24 near zero modes with the representation content of Eq. (4). Symbols and axes as in Fig. 1.

capsid. The capsids in our examples are not invariant under inversion, hence the symmetry group is $\mathcal{I}$, and the irreducible representations are all of type "+." Discarding the two weakest bonds (which are both edge crossing) leads to $k=4$ and 9 edge-crossing bonds per edge. In agreement with $2(3n-k)-1=9$ we then observe 24 nontrivial zero modes. Restoring the weak long-range arms which stretch from the (green) C chains has the crucial effect of lifting the 24 zero modes to produce a low-frequency plateau. The $T=7l$ Hong-Kong 97 bacteriophage (PDB: 2fte) has a stable capsid too. Keeping all but the six weakest bonds (only one of them is edge crossing in Fig. 4) leads to $k=12$ and 17 edge crossing bonds per edge, in accordance with $2(3n-k)-1=17$, so that there are 24 nontrivial zero modes. The weak bonds serve to lift these zero modes, and the spectrum indeed exhibits again a very low-frequency plateau of 24 modes (see Fig. 4). A variety of other capsids with similar spectral signatures are presented in Ref. [26].

IV. CONCLUSION AND DISCUSSION

We have analyzed a model of icosahedral virus capsids which incorporates areas of varying flexibility. We have given a simple mathematical argument which explains the appearance of a low-frequency plateau of 24 states in the vibrational spectrum of Caspar-Klug viruses. Our top-down approach should be viewed as a first step in bridging the gap with the few molecular dynamics analyses published on viral capsids, and our results should be viewed as a remarkable property of the lowest order approximation to virus capsid vibrations.

We wish to stress that if the amplitudes of the very low frequency modes were known, they could help decide whether they trigger instabilities and some of the conformational changes observed experimentally. Such changes play a fundamental role in the function of viruses, and a fundamental explanation of their origin is still lacking. It is moreover widely believed that the only motions that change the capsid surface, and hence influence significantly the interactions of the virus with the environment, are those global large amplitude, slow motions. Hence the importance of developing models that pin them down as accurately as possible. Another application of their knowledge is to provide experimenters with systematics on how to tune laser pulses in the near infrared to produce damage on viral capsids by forced resonance, an innovative technique developed recently [27].

We believe that our model can be developed beyond the crude approximation we have used in our paper, and would then greatly benefit from getting data on the absolute strength of the bonds, which would provide us with quantitative information on low-frequency motions. In order to test our results, however, improved techniques are required as there are, at present, no experimental setups which could reliably and directly check our frequencies predictions (as the frequencies we target are too low for commonly used techniques).

ACKNOWLEDGMENTS

The work of F.E. was partly supported by IISN-Belgium (convention 4.4511.06 and convention 4.4505.86), and by the Belgian Federal Science Policy Office through the Interuniversity Attraction Pole P VI/11. The work of K.P. was partly supported by VIDI Grant No. 016.069.313 from the Dutch Organization for Scientific Research (NWO). Both thank the Department of Mathematical Sciences at Durham University for its warm hospitality.

[1] J. C. Smith, *Structure and Dynamics of Biomolecules* (Oxford University Press, Oxford, 2000), pp. 161–180.
[2] J. A. McCammon, B. R. Gelin, M. Karplus, and P. Wolynes, Nature (London) **262**, 325 (1976).
[3] T. Noguti and N. Go, Nature (London) **296**, 776 (1982).
[4] B. R. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **80**, 6571 (1983).
[5] N. Go, T. Noguti, and T. Nishikawa, Proc. Natl. Acad. Sci. U.S.A. **80**, 3696 (1983).
[6] M. Levitt, C. Sander, and P. Stern, Int. J. Quantum Chem. **10**, 181 (1983).
[7] R. Harrison, Biopolymers **23**, 2943 (1984).
[8] R. Elber and M. Karplus, Science **235**, 318 (1987).
[9] H. Frauenfelder, F. Parak, and R. D. Young, Annu. Rev. Biophys. Biophys. Chem. **17**, 451 (1988).
[10] M. K. Hong *et al.*, Biophys. J. **58**, 429 (1990).
[11] B. R. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **82**, 4995 (1985).
[12] J. F. Gibrat and N. Go, Proteins **8**, 258 (1990).
[13] O. Marques and Y. H. Sanejouand, Proteins **23**, 557 (1995).
[14] L. Mouawad and D. Perahia, J. Mol. Biol. **258**, 393 (1996).
[15] V. Alexandrov, U. Lehnert, N. Echols, D. Milburn, D. Engelman, and M. Gerstein, Protein Sci. **14**, 633 (2005).
[16] T. Simonson and D. Perahia, Biophys. J. **61**, 427 (1992).
[17] F. Tama and C. L. Brooks, J. Mol. Biol. **318**, 733 (2002).
[18] H. W. T. v. Vlijmen and M. Karplus, J. Mol. Biol. **350**, 528 (2005).
[19] A. Rader, D. Vlad, and I. Bahar, Structure (London) **13**, 413 (2005).
[20] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[21] J. I. Jeong, Y. Jang, and M. K. Kim, J. Mol. Graphics Modell. **24**, 296 (2006); Nucleic Acids Res. **4**, 382 (2006); Nucleic Acids Res. **34**, W57 (2006).
[22] D. Caspar and A. Klug, Cold Spring Harbor Symp. Quant. Biol. **27**, 1 (1962).
[23] C. Shepherd, I. Borelli, G. Lander, P. Natarajan, V. Siddavanahalli, C. Bajaj, J. Johnson, C. I. Brooks, and V. Reddy, Nucleic Acids Res. **34**, D386 (2006).
[24] K. Peeters and A. Taormina, Comput. Math. Methods Med. **9**, 211 (2008).
[25] M. Widom, J. Lidmar, and D. R. Nelson, Phys. Rev. E **76**, 031911 (2007).
[26] K. Peeters and A. Taormina, e-print arXiv:0806.1029
[27] K. T. Tsen, S.-W. D. Tsen, O. F. Sankey, and J. G. Kiang, J. Phys.: Condens. Matter **19**, 472201 (2007).
[28] $n=h^2+hk+k^2$ for h and k integer and relatively prime.
[29] For the simpler icosahedrically symmetric C_{60}, our approximation would involve 60 atoms connected by 90 bonds. Our prescription for counting the number of nontrivial zero modes of this system suggests a plateau of 84 nontrivial low-frequency modes, rather than 24. However, it is unlikely that this result can be used to argue that, in the presence of a realistic potential, one should observe a gap in the spectrum. The 84 zero modes now span half of the total number of states, which is a much larger fraction than the zero modes of a $T=n$ capsid ($\frac{1}{6n}$), especially for $n>1$. As each subsequent zero mode will be lifted by a small amount, perturbation theory is likely to become invalid as one approaches the end of the plateau.
[30] In Ref. [24] the point-mass coordinates were projected radially onto icosahedron faces, which blurs stability issues.